NPS-51MR720301A

NAVAL POSTGRADUATE SCHOOL

Monterey, California

REGRESSION METHODS IN FIVE-DAY MEAN HEIGHT PREDICTION AT 500MB USING MEAN NIMBUS II COMPOSITED FOR IDENTICAL PERIODS IN THE UPDATE PROCEDURE

by

Frank L. Martin

31 March 1972

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral A. S. Goodfellow
Superintendent

M. U. Clauser
Provost

ABSTRACT

A series of 15 successive 5-day mean 500mb height fields in gridprint form covering the quasi-hemispheric area between 14.8 - 70.2N latitude was used in this experiment. The time period extended between 15 May 1966 through 28 July 1966. The stepwise regression procedure was employed to derive statistical estimator equations for each mean map field $Z_5(I,J,t+1)$ in terms of the preceding field data $Z_5(I,J,t)$ as the primary predictor, (t+1 = 2, ..., 15). As additional predictors, the 5-day mean composited gridpoint values of the NIMBUS II MRIR equivalent black body temperatures in the water-vapor channel (denoted T_1) and the window channel (denoted T_2) were introduced for both analyses times t and t+1. The five-predictor regression equations thus developed for $Z_5(I,J,t+1)$ proved to have substantial statistical significance, both from the standpoint of the specification and of prediction of $Z_5(I,J,t+2)$ valid five days later (t+2 = 3, ..., 15). In the latter connection, the prognosis of $Z_5(t+2)$ was significantly improved when all five predictors were included, as compared with the case when the data of either radiometric channel $T_1(I,J)$ or $T_2(I,J)$ at times t+1, t+2 was deleted from the prediction equations.

TABLE OF CONTENTS

1. Introduction

The objective of this study was to determine the usefulness of introducing satellite radiometric data in developing improved regression equations for the prognosis of the 500mb contour height field. The measure of the improvement is based upon a comparison with the skill indicated by use of a simple lag-correlation analysis between $Z_5(t+1)$ and $Z_5(t)$ over all grid points of the prediction area.

A series of 15 successive 5-day mean 500mb Northern Hemisphere analyses were provided on magnetic tape through the kind auspices of the Extended Forecast Section of the U.S. Weather Bureau. These analyses ranged over the successive non-overlapping periods 15 - 19 May 1966 (denoted $Z_5(1)$), and on in order to the final analysis of 24 - 28 July 1966 [denoted $Z_5(t=15)$]. This set of 15 analyses provided 15 fields $Z_5(I,J,t)$, $(t=1, \ldots, 15)$, where I,J are the geographic grid-point identifiers.

The rationale of employing 5-day mean analyses in this experiment was based upon the consideration that the composited 5-day mean radiometric data fields of NIMBUS II over identical time periods would be obtained with as closely comparable a degree of feature-smoothing as is involved in the Extended Forecast Section's analysis procedure for deriving $Z_5(I,J,t)$.

The 5-day mean radiometric fields employed were the grid-point composited values of equivalent black body temperature T_{BB} in the medium resolution infrared radiometer (MRIR) channels 1 and 2 of NIMBUS II (c.f., NIMBUS II User's Guide, 1966). For this satellite, these two channels were also called the water-vapor channel (6.4 - 6.9μm) and window-channel (10-11μm), respectively. With this understanding, these fields were denoted $T_1(I,J,t)$ and $T_2(I,J,t)$, respectively. The full lifetime of 75 days of observational data of the NIMBUS II MRIR scanning radiometer was available for this experiment,

thus accounting for the duration of the test and the number of analyses involved (t=1, ..., 15). The composited radiometric data were provided in hard-copy form through the generous cooperation of the National Space Science Data Center.

Other radiometric fields from NIMBUS II could have been included in the regression study. However, in an earlier study (Martin, 1969) showed that channel 4, which sensed in the spectral range 5-30μm, was virtually redundant with channel 2 since there existed a spatial correlation coefficient $r(T_2, T_4) \approx .98$. Furthermore, the choice of the two infrared channels used here seems to afford the best possibility of continuity for application of the method developed to future generations of polar orbiting satellites planned for the 1970's.

2. The data.

a. The time series.

Three separate time-series of analysed 5-day mean map fields were processed into similar grid-map arrays for each of the identical analysis times t (t=1,...., 15):

t = 1, 00GMT, 20 May 1966; t = 2, 25 May 1966

t = 3, 30 May 1966; t = 4, 4 June 1966

.....

.....

t = 9, 29 June 1966; t =10, 4 July 1966

t =11, 9 July 1966; t =12, 14 July 1966

t =13, 19 July 1966; t =14, 24 July 1966

t =15, 00GMT, 29 July 1966.

Each analysis time specified represents the averaging period over the just-preceding five days (e.g. t=1 denotes an average over 15-19 May, which may be characterized as the date-time analysis for 00GMT, 20 May 1966).

b. Geographic extent of the original fields.

(1) The 5-day mean fields of 500mb geopotential height were initially in the usual NMC polar stereographic map-projection coordinates (i,j).

(2) The 5-day mean fields $T_1(I,J,t)$ were quasi-global spanning the region 70S-70N in Mercator projection coordinates (I,J) at 5^o longitude intervals.

(3) The field $T_2(I,J,t)$ covered the same region as $T_1(I,J,t)$. The Mercator grid points (I,J) were converted into longitude λ and latitude φ through the formulas

$$\lambda = 30^o + (I - 1.)5^o \qquad I = 01,.....,73 \qquad (1)$$

$$\ln\left(\frac{1+\sin\varphi}{\cos\varphi}\right) = (21. - J)(5\pi/180) \qquad J = 01,.....,41 \qquad (2)$$

Eq. (1) gives longitude λ in degrees east of Greenwich, while (2) gives φ (in radians) on a non-linear scale which is defined in terms of J in Table 1. In this table, equal magnitudes $|J-21.|$ correspond to equal latitudes φ relative to the equator. This is indicated in Table 1 by coupling the symmetric values J, 42-J under the column-head "J-value," with identical φ-values north or south in the adjacent column.

An example of the Mercator mapping of $T_2(I,J,t=6)$ is shown in Fig. 1. In this figure, a rather heavy degree of smoothing has been applied to the original data over the 73 by 41 array but the smoothing merely serves to highlight the large-scale global map features, and is otherwise irrelevant to the study outlined here.

In this study, each radiometric field was encoded by rows in the form indicated by

$$T_2\ [I,J,t],\ I = 1,....,\ 72;\ J = 41,....,01$$
$$t = 1,....,\ 15.$$

As indicated earlier, no spatial smoothing was performed on any radiometric data array, since the value recorded at any grid point (I,J) was already composited of a large number of scan-spots within the 5-degree longitude "square" surrounding (I,J). All fields $T_1(I,J,t)$, $T_2(I,J,t)$ were converted to IBM card format in the manner just suggested.

c. Collection of data-arrays in compatible Mercator format.

Since $Z_5(i,j,t)$ is considered to be a conventional form of analysis, while $T_1(I,J,t)$, $T_2(I,J,t)$ are experimental for the purposes of

TABLE 1. Conversion from J-value on Mercator grid-map to corresponding latitude, (north or south) using Eq.(2).

J-value	Lat.(deg.) by Eq.2	J-value	Lat.(deg.) by Eq.2
01,41	70.2(N,S)	12,30	41.0 (N,S)
02,40	68.4	13,29	37.1
03,39	66.5	14,28	33.0
04,38	64.4	15,27	28.7
05,37	62.2	16,26	24.2
06,36	59.8	17,25	19.6
07,35	57.2	18,24	14.8
08,34	54.3	19,23	9.9
09,33	51.3	20,22	5.0
10,32	48.1	21	0
11,31	44.6		

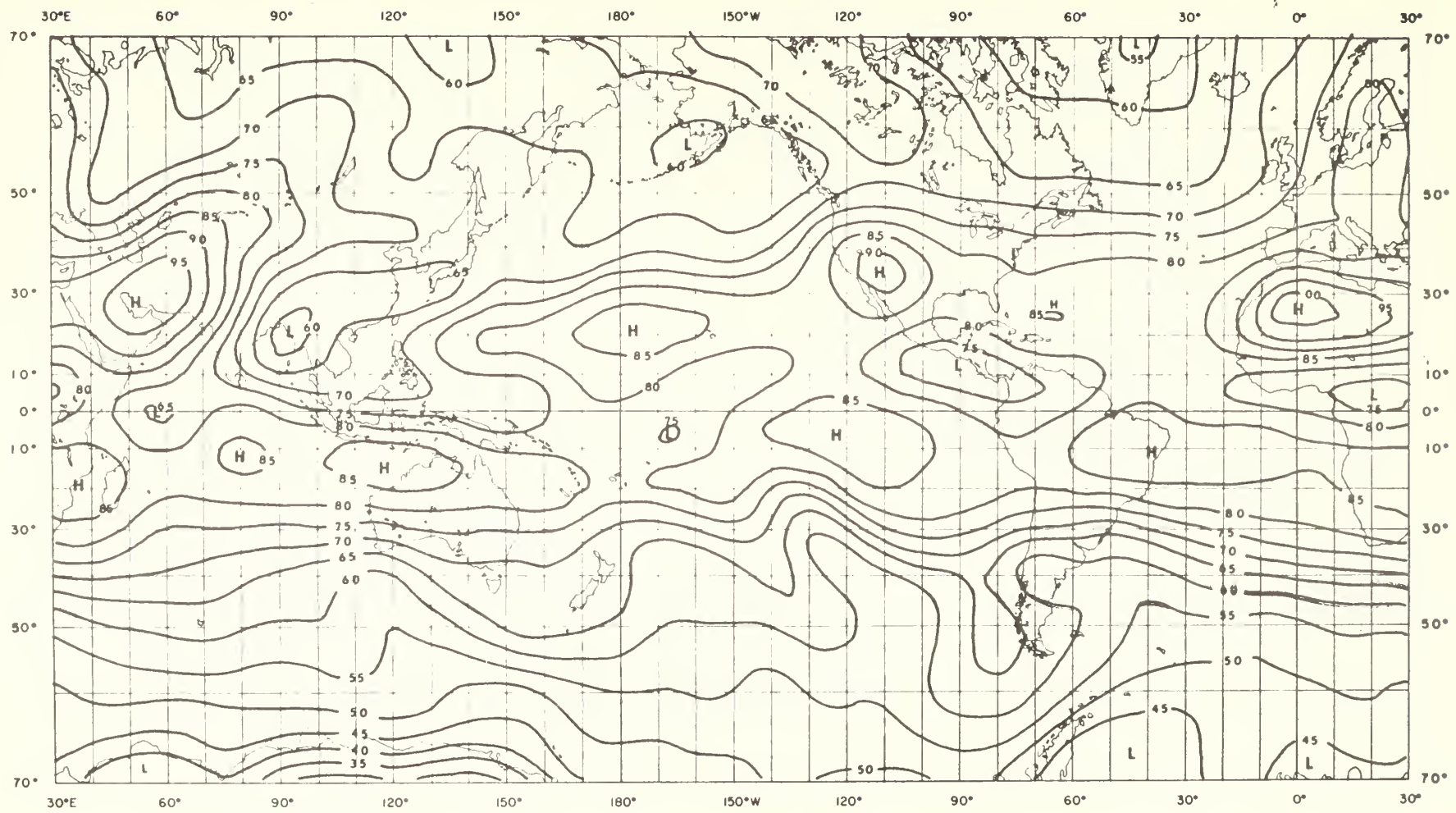

FIG. 1. An example of Mercator mapping of T_{BB} for the window-channel composited field for the 5-day period 9-13 June 1966, inclusive (i.e., for t= 06). Isolines of T_{BB} are labeled with the last two digits of Kelvin temperature: 70 = 270K.

this study, the former has been converted to the Mercator-mapping described in equations (1) and (2). The geometric relationships involved in the transformation from polar stereographic to Mercator coordinates are

$$\tan (\lambda - 10^{o}) = \frac{j-32}{i-32} \qquad (3)$$

and

$$\sin \varphi = 973.71 - [(i-32.)^2 + (j-32.)^2] / \{[973.71 + (i-32.)^2 + (j-32.)^2]\} \qquad (4)$$

Equations 3 and 4 make it possible to compute $Z_5(I,J)$ at whole multiples of 5^{o} longitude and at values J which are compatible with φ of Eq. (2).

Because the conventional NMC polar stereographic chart has an equatorward boundary which averages close to latitude 13N, only linear interpolation of $Z_5(I,J)$ along J = 18 (corresponding to 14.8N) could be performed. However, for all $J \leq 17$, (i.e. $\varphi \geq 19.6$N) all other points (I,J) were sufficiently well removed from the NMC boundary that Bessel's central difference interpolation scheme [p. 252, Haltiner (1971)] could be applied at all other Mercator points. An example of the NMC map field at t = 6, after conversion to the Mercator projection (1),(2), is shown in Fig. 2.

It should be noted that the resulting conversion to Mercator coordinates leaves us with contour-height data defined only on rows J=18, 17,...., 01 and for the 72 columns I=01,...., 72. Thus, for purposes of point-by-point regression, all radiometric data points in $T_1(I,J,t)$, $T_2(I,J,t)$ with $J > 18$ were stripped from their card decks. For each time t, the Mercator-listed fields

$$T_1(I,J,t),\ T_2(I,J,t),\ Z_5(I,J,t) \quad J = 18,...., 01,\ I = 01,...., 72$$
$$t = 1,...., 15$$

were arranged in sequential order with respect to I and J, taking care not to break the sequence within any field. In the form just specified, each field T_1, T_2, Z_5 consisted of 1296 geographically ordered data elements valid at time t.

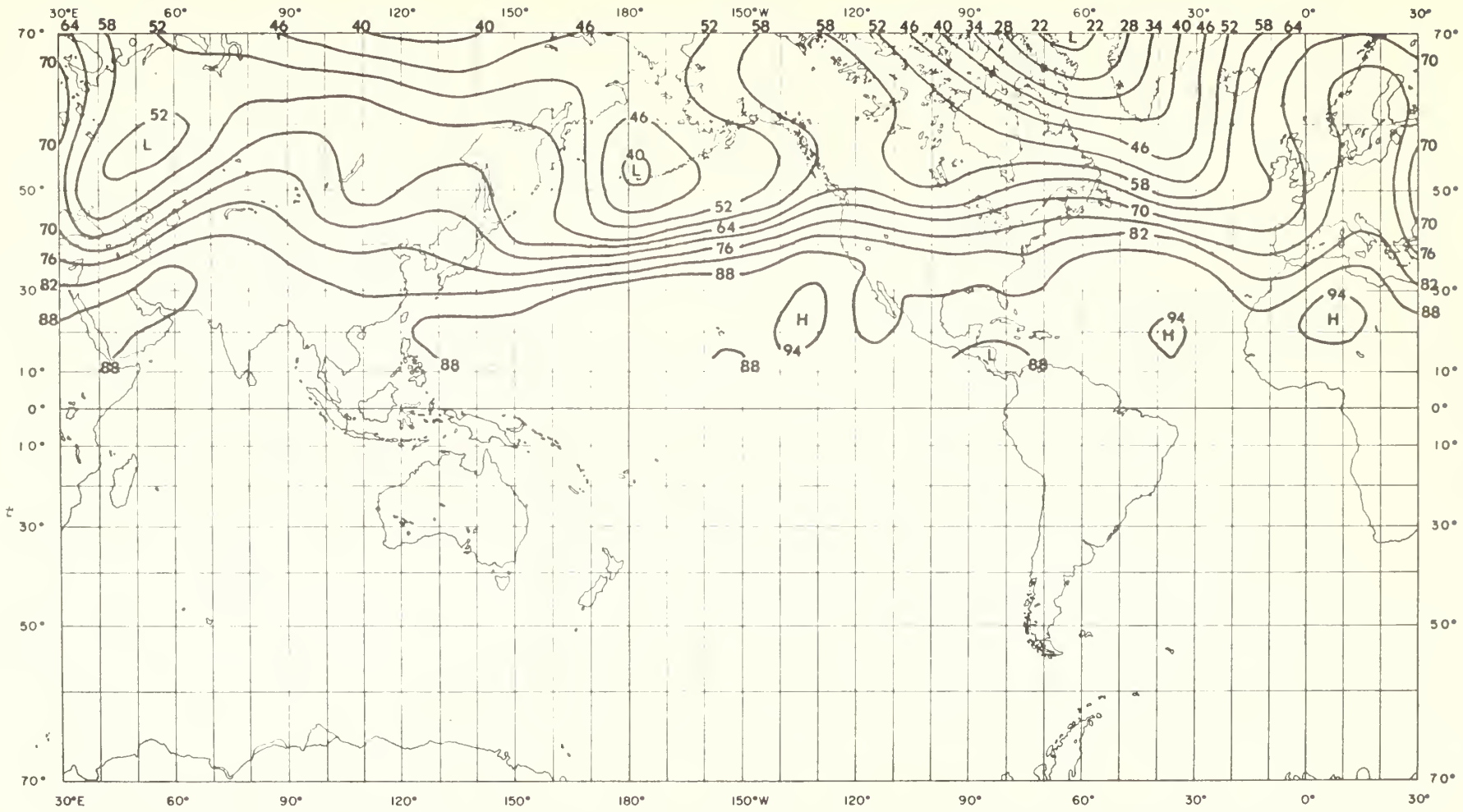

FIG. 2. The 5-day mean 500-mb contour analysis for the period 9-13 June 1966, inclusive (t= 06), in Mercator coordinates between 14.8N and 70.2N. The isopleth 94 represents 594 decameters.

A Fortran program was devised to sort all data elements at (I,J) to produce a new punched card deck having the rearranged data array A on each card:

Data Array A at (I,J):

$T_1(I,J,t)$, $T_1(I,J,t+1)$, $T_2(I,J,t)$, $T_2(I,J,t+1)$, $Z_5(I,J,t)$, $Z_5(I,J,t+1)$

$I = 1,\ldots, 72,$ $J = 18,\ldots, 01.$

Each data array of the form A now applies spatially over each of the 18 x 72 = 1296 Mercator gridded points for the successive times (t, t+1). This array by paired times will be considered valid at time t+1 in the statistical regression experiment to be performed in Section 3.

3. The statistical experiments on the dependent-data arrays.

For each data array A spanning the entire 1296 grid point map field (I,J) at time t+1, three separate tests were performed, as described below. In Test 1, a one-predictor specification equation was generated in the form

$$Z_5(t+1) = A_o + A_5 Z_5(t) \tag{5}$$

whereas in Tests 2 and 3, respectively, three- and five-predictor specification equations were generated from the same grid-sample data base. Thus, the Test 2 regression equation was of form

$$Z_5(t+1) = B_o + B_3 T_2(t) + B_4 T_2(t+1) + B_5 Z_5(t) \tag{6}$$

while the Test 3 equation was developed in the form

$$Z_5(t+1) = C_o + C_1 T_1(t) + C_2 T_1(t+1) + C_3 T_2(t) + C_4 T_2(t+1) + C_5 Z_5(t) \tag{7}$$

The regression equations of form (5), (6), (7) were derived simply by using the delete-variable option of the BIMED02R stepwise regression program [designed after Dixon (1966)] on file in the program library of the W. R. Church Computer Center of the Naval Postgraduate School.

In the progressive application of the program to develop the regression equations (5), (6), (7) at each time t+1, the significance of the kth variable selected could be assessed using the F_k-statistic upon entry (an output of the program at each step). This statistic may be defined as

$$F_k(1,n-k-1)=\frac{[\%\ \text{cum.expl.var., step k}]-[\%\ \text{cum.expl.var., step (k-1)}]}{[\%\ \text{unexplained variance at step k}]} \qquad (8)$$

Here, n is the number of degrees of freedom, taken equal to the sample size, n = 1296. Miller's criterion (1962) for significance of the regression equation at or above the 95% confidence level is that each predictor selected yield an $F_k(1,n-k-1)$ in excess of a critical value at the kth step defined in

$$F^{(c)}_{.05/(P-k+1)}(1,n-k-1)$$

Such critical values may be obtained from tables of F-values for a 95% confidence estimate at each step. The values for Tests 1 and 2 are bounded by those obtained in the five-predictor case (P=5), for which case $F_k^{(c)}$ was found to have the following critical values at step k, (k=1,...,5):

Step k =	1	2	3	4	5
$F^{(c)}_{\alpha/P-k+1}$ =	6.63	6.24	5.76	5.02	3.84

In each regression equation developed over the entire time span whether by Test 1, 2 or 3, all variables used in equations 5, 6, 7 were found to be significant, at least at the 95% confidence level. A case-by-case listing of the coefficients of the 14 regression equations by each test is given in Appendix Table 5. For the remainder of this section, a brief summary of the statistics resulting from the use of these equations will suffice. These summarized statistics are displayed in Table 2.

TABLE 2. Summary of the results of the specification Tests 1, 2, 3. The time mean values of the coefficients of the variable

$$Z_5(t),\ T_2(t+1),\ T_2(t),\ T_1(t+1),\ T_1(t)$$

are shown along with their standard deviations (S.D.). In addition, $\overline{R^2}$ the mean fractional explained variance, averaged over each time series and by each test is listed.

Test Series	Mean-value of coefficients of predictors of Eqs.(5,6,7)						
	$Z_5(t)$	$T_2(t+1)$	$T_2(t)$	$T_1(t+1)$	$T_1(t)$	Const.term (gpm).	R^2
1	.87456	0	0	0	0	93.560	.8063
(S.D.)	(.06116)	0	0	0	0	(46.535)	
2	.78742	4.46333	-2.32201	0	0	-432.644	.8413
(S.D.)	(.07622)	(1.2540)	(1.28331)	0	0	(239.823)	
3	.78316	6.65738	-3.75340	-8.47452	5.81224	-7.978	.8533
(S.D.)	(.08019)	(1.59336)	(1.38510)	(3.39685)	(3.07409)	(206.807)	

The significance in each test of the multiple correlation coefficient R and its counterpart R^2, is that the latter represents the fractional explained variance of $Z_5(t+1)$ by whichever of the regression equations 5,6 or 7 is under consideration.

The general significance of the results summarized in Table 2, may be described by means of an F-value characterizing Test 1, an F-value for the difference between Tests 2 and 1, and a final F-value for the difference between Tests 3 and 2. The first of these F-statistics results from Eq.(8) using k=1, while the other two F-values may be expressed [after Anderson (1960)], in the form

$$F_{5-3} = \frac{(R_5^2 - R_3^2)/2}{(1 - R_5^2)/1290} \quad \text{and} \quad F_{3-1} = \frac{(R_3^2 - R_1^2)/2}{(1 - R_3^2)/1292}$$

Here, the subscripts 5, 3, 1 signify the number of predictors in the regression equation under consideration.

The three F-values are

$$F_1 = 5{,}386.4 \qquad F_{3-1} = 101.8 \qquad F_{5-3} = 52.8$$

each of which is highly significant when compared with the critical F-level required at the 99% confidence level.

To sum up, there is a high degree of specification skill in the stepwise addition of each variable, or set of variables, involved in proceeding from Test 1 through Test 3.

4. Application of the regression equations to independent data

In arriving at the precise form of equations 5, 6, 7 at each specification time t+1 = 2, ..., 15, matrix equations of form

$$Z_5(t+1) = (C_o, C_1, C_2, C_3, C_4, C_5)_{t+1} \begin{pmatrix} 1 \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{pmatrix}_{t+1} \qquad (9)$$

have been derived using the predictor-variables valid at t+1:

$X_1 = T_1(t)$, $X_2 = T_1(t+1)$, $X_3 = T_2(t)$, $X_4 = T_2(t+1)$, $X_5 = Z_5(t)$

and with $Z_5(t+1)$ as the predictand. Eq.(9) affords a summary of each of three tests if the coefficients C_i are properly interpreted for each test (see Appendix Table 5). The coefficient matrix deduced for the data array at time t+1 was then applied to the data array at time t+2 to generate an estimated field $\tilde{Z}_5(t+2)$ from the matrix equation

$$\tilde{Z}_5(t+2) = (C_o, C_1, \ldots, C_5)_{t+1} \begin{pmatrix} 1 \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{pmatrix}_{t+2} \qquad (10)$$

The prognostic $\tilde{Z}_5$-field of (10) was next correlated with the verifying $Z_5(t+2)$ giving a resultant multiple correlation coefficient $R(Z_5, \tilde{Z}_5)_{t+2}$ and a standard error of estimate σ_e of the predicted field given by the standard formula

$$\sigma_e(t+2) \doteq \sigma_Z(t+2)[1 - R_k^2(t+2)]^{1/2} \qquad (11)$$

The estimated field $\tilde{Z}_5$ was generated from the data-array A at time (t+2) using an option of the statistical program BIMED 02R, which also computed the resulting statistical parameters R_1^2, R_3^2, R_5^2 together with the associated standard errors σ_e. The results using (10) iteratively on all possible independent data times t+2 = 3,....,15 and for each test are displayed in Table 3.

For comparison purposes, the corresponding dependent sample statistics R_1^2, R_3^2, R_5^2 and standard deviations based upon the results of the specification equations 5, 6, 7 are also listed for each test method and validity time t+1 = 2,...,15. The results for these dependent samples are those which were time-averaged in the preliminary statistical analysis of Section 3.

It is to be noted that in any time-line (t+2), except the first line of Table 3, there will be both a dependent-data test by Eqs.

TABLE 3. Results of the prognostic equation of form (10) for $\tilde{Z}_5(t+2)$ applied to the independent data at times t+2 = 3,....., 15, by each of Tests 1, 2 and 3. For comparison, the results of the specifications on dependent data at times t+1 = 2, ..., 15 are shown. The statistic R^2 is the fractional explained variance in each case considered. Values of the standard error (in gpm) after each specification and prognosis are listed based upon Eq.(11).

Analysis time	Test 1 R_1^2 Dep. data	Test 1 R_1^2 Indep. data	Test 2 R_3^2 Dep. data	Test 2 R_3^2 Indep. data	Test 3 R_5^2 Dep. data	Test 3 R_5^2 Indep. data	Standard error of estimate (gpm) Test 1 Dep.	Test 1 Indep.	Test 2 Dep.	Test 2 Indep.	Test 3 Dep.	Test 3 Indep.
t+1 = 2	.8574	--	.8856	--	.9084	--	71.315	--	63.918	--	57.243	--
3	.7743	.7743	.8029	.8028	.8257	.8257	76.023	76.023	71.105	71.108	67.909	67.883
4	.7441	.7441	.8005	.7993	.8090	.8027	89.500	89.500	79.047	79.226	77.404	78.547
5	.7954	.7954	.8395	.8382	.8465	.8444	77.747	77.747	68.901	69.141	67.440	67.799
6	.8587	.8587	.8847	.8801	.8999	.8944	65.518	65.518	59.221	60.336	55.233	56.628
7	.8772	.8772	.8925	.8896	.9001	.8942	62.336	62.336	58.352	59.031	56.298	57.856
8	.8183	.8183	.8652	.8581	.8688	.8604	69.349	69.349	59.784	61.299	59.024	60.799
9	.8723	.8723	.8908	.8787	.8970	.8832	55.584	55.584	51.428	54.171	49.994	53.146
10	.7946	.7946	.8331	.8290	.8427	.8381	70.103	70.103	63.236	63.701	61.443	62.245
11	.8196	.8196	.8390	.8375	.8464	.8425	63.361	63.361	59.915	60.140	58.502	59.207
12	.8550	.8550	.8655	.8586	.8787	.8616	58.918	58.918	56.749	58.172	53.966	57.549
13	.8150	.8150	.8428	.8332	.8579	.8458	62.945	62.945	58.067	59.771	55.265	57.465
14	.6711	.6711	.7573	.7370	.7725	.7465	80.912	80.912	69.557	72.349	67.392	71.031
15	.7350	.7350	.7805	.7384	.7925	.7488	72.034	72.034	65.613	71.578	63.853	70.134
Time-mean	.8063	.8063	.8413	.8293	.8533	.8376	69.206	69.041	63.632	64.993	61.112	63.493

15

(5),(6),(7) and a test on the independent data, when the results of Eq.(10) are applied to the data at time (t+2).

Time-averages of each column are also presented in Table 3. Since the standard errors are root mean square values, the time-averaged values of σ_e were obtained by squaring the individual-time values, and averaging the squared quantities prior to recomputing the root mean squares.

5. Analysis of the statistical results.

Examination of the sequence of values R_1^2, R_3^2, R_5^2 first under the headings "dependent-data" reveals the same conclusion as in Section 3; that is, the percentage explained variance increases monotonically as more variables are added to the specification equation (9).

A similar result obtains when a comparison is made of the independent-data verification-results R_1^2, R_3^2, R_5^2, by the different tests at time (t+2). It is found, as before, that

$$R_5^2(t+2) > R_3^2(t+2) > R_1^2(t+2)$$

although the successive differences in R_k^2 are not quite as large as with the dependent-data differences for the same time.

A test of the sensitivity of the prediction coefficient matrix is considered next. Our remarks will be directed specifically to Test 3, although similar remarks apply to Tests 1 and 2. If the success of the prognosis was strongly dependent upon the coefficient matrix valid at time (t+1), one would expect a greater coherence between the

dependent $R_5^2(t+1)$ and the ensuing (independent) $R_5^2(t+2)$. In fact, a comparison of R_5^2 from the dependent to independent-data columns indicates that there are four cases when $R_5^2(t+2)$ exceeds $R_5^2(t+1)$; in two cases there is only a very slight change, and in the other seven cases, there is a sizeable decrease in R^2 in stepping timewise from (t+1) to (t+2). On the other hand, there is always a slight shrinkage at the same time t+2 in proceeding from the dependent-data to the independent-data samples.

This combination of results indicates that the successive-time regression coefficients (C_o, C_1,....,C_5) frequently did not pass outside of the standard error involved in their determination. For example, in the Test 1 cases, $R_1^2(t+2)$ based upon both the dependent-data specification and the independent data-verification which used the coefficient matrix at time (t+1), were identical to four significant digits.

A separate test of the inertia of the coefficient-matrix was conducted using three-period pooled specification equations of form (7) to determine the coefficient matrix for prognosis at time (t+2). For example, the pooled sample of 3988 data elements from times 2,3,4 were used to provide a coefficient matrix to test the prognosis $\tilde{Z}_5(t+2=5)$, and so on iteratively throughout the data period to t+2=15. In no case was there any significant variation in explained variance of the prognosis arising from the use of the three-period coefficient matrix

sets as compared with the single-period coefficient matrix from time t+1 = 4,....,15. Most of the success in the prognosis procedure must, therefore, be due to the quality of the dependent-data sample specification, that is in the sample linear correlations, and partial correlations which relate $Z_5(t+2)$ to the five variables $X_1,....,X_5$ identified just below Eq.(9), but now considered valid at time t+2.

In spite of the conclusion that the linear regression prognosis model (10) has coefficients which are temporally smoothed, it is still of considerable value to determine the relative usefulness in the prognostic procedure of the various prediction equations identified as relevant to Tests 1, 2, 3. In order to demonstrate the character of the results, the additional fractional explained variance by Test 2 compared to 1, by Test 3 relative to 2 and Test 3 relative to 1, in the prognosis mode is summarized in Table 4. Since the corresponding dependent-sample statistics are to be considered optimal at the same time as the prognosis verified,the identical-time dependent-data statistics are listed. It should be recalled, however, that the prognosis-mode associated with the independent-data sample employs the coefficient matrix determined at time t+1.

TABLE 4. Added percentage explained variance by Test 2 relative to Test 1, by Test 3 relative to 2, and Test 3 relative to 1. The comparative results are shown corresponding to the same time-line (t+2) in Table 3 by means of the parameters (% added explained variance) $100(R_3^2 - R_1^2)$, etc. Time-means of $100(R_3^2 - R_1^2)$, etc., are taken over the 13 times t+2=3,...,15.

	Test (2-1) $100(R_3^2 - R_1^2)$		Test (3-2) $100(R_5^2 - R_3^2)$		Test (3-1) $100(R_5^2 - R_1^2)$	
	Dep.	Indep.	Dep.	Indep.	Dep.	Indep.
MIN	1.05 (t+2 = 12)	0.36	0.36 (t+2 = 8)	0.23	2.37 (t+2 = 12)	0.66
MAX	8.62 (t+2 = 12)	6.59	2.28 (t+2 = 3)	2.29	10.14 (t+2 = 14)	7.54
MEAN	3.58	2.30	1.13	0.83	4.71	3.13

In Table 4 is shown the minimum and maximum of percentage explained variance by proceeding from the use of Eqs. (5) to (7) in the time series of data-arrays A. The results are shown both for the independent data-sample at (t+2), as well as the added explained variance using the series of specification equations valid at the same time. In the case of minimum and maximum added variance, the period number (t+2) observed from Table 3 has been noted under the corresponding event in Table 4.

It is apparent that the use of the five-predictor equation (7) and/or (10) yields consistent additional explained variance in the prognostic mode over both one- and three-predictor equations.

6. Some remarks concerning the regression-update procedure.

Examination of Tables 3 and 4 indicates the general nature of criteria for a "successful" prognosis by use of Eq.(10). Temporarily, the term "successful" will be regarded as $R_5^2(t+2)$ upon prognosis as exceeding 84% of the percentage explained variance.

Guided by Table 4 but with specific reference to the case-by-case listings in Table 3, a typical profile for a successful prognosis may be sketched in broad brush form. The prime requirement appears to be that $R_1^2(t+2) \geq 0.795$ upon specification.

At the same time, the added percentage explained variance upon specification when the T_2-variables are included should be of the order 3 to 5%, depending upon how much $R_1^2(t+2)$ exceeded 0.795. The added explained variance upon inclusion of the T_1-variables does not appear to be so crucial using the data of NIMBUS II, although some additional explained variance ranging from 0.5 to 1.0% seems to be a reasonable expectation (Table 4).

Some examples of relatively inefficient prognoses occurred at times t+2 = 4, 14, 15 (Table 3) when the values of the simple lag-correlation coefficient-squared, $R_1^2 = R^2[Z_5(t+2), Z_5(t+1)]$, were 0.7441, 0.6711, 0.7350, respectively. One does not know, a priori, at time t+2 what this value of $R_1^2(t+2)$ will prove to be. However, one method of testing for a tentative value of $R_1^2(t+2)$ would be to compute $\tilde{Z}_5(t+2)$, with the update radiative fields in hand, and then computing the simulated $\tilde{R}_1^2(t+2)$. An examination of Test 1 from

Table 3 suggests that the proper trend in $R_1^2(t+2)$ would thus be indicated. For the most part, however, the results upon prognosis (independent data) in Table 3 with the exception of t+2 = 14, 15, were encouraging.

Two examples of prognoses and their comparisons with analyses at the same verifying times are given in Section 7.

7. Some graphical prognoses and their verifications.

In this section, two experiments are described based upon the use of the five-predictor Eq.(10) derived from the coefficient-array specified at time t+1 and applied to a prognosis verifying at time t+2. The two prediction equations generated correspond to the use of the coefficient-matrices

$$(C_o, C_1, \cdots, C_5)_{t+1=5} \; ; (C_o, C_1, \cdots, C_5)_{t+1=11}$$

taken from Table 5. These were then applied to the data arrays at t = 06 and 12, respectively for the computation of

$$\tilde{Z}_5(t+2=06) \quad \text{and} \quad \tilde{Z}_5(t+2=12)$$

Contours were drawn of intervals of 60.0 gpm, over the map fields shown in Fig. 3 and 6. These two maps depict the prognoses for

$$\tilde{Z}_5(I,J,t=06) \quad \text{and} \quad \tilde{Z}_5(I,J,t=12)$$

are
and/to be verified against the actual contour analyses shown in Figs. 2 and 5 respectively at times t=06 and t=12. The corresponding error fields $(\tilde{Z} - Z)$ are shown in Figs. 4 and 7.

In the case of $\tilde{Z}_5(t=06)$, Table 3 shows that fractional explained variance was .8944 for the independent (verifying) data. This

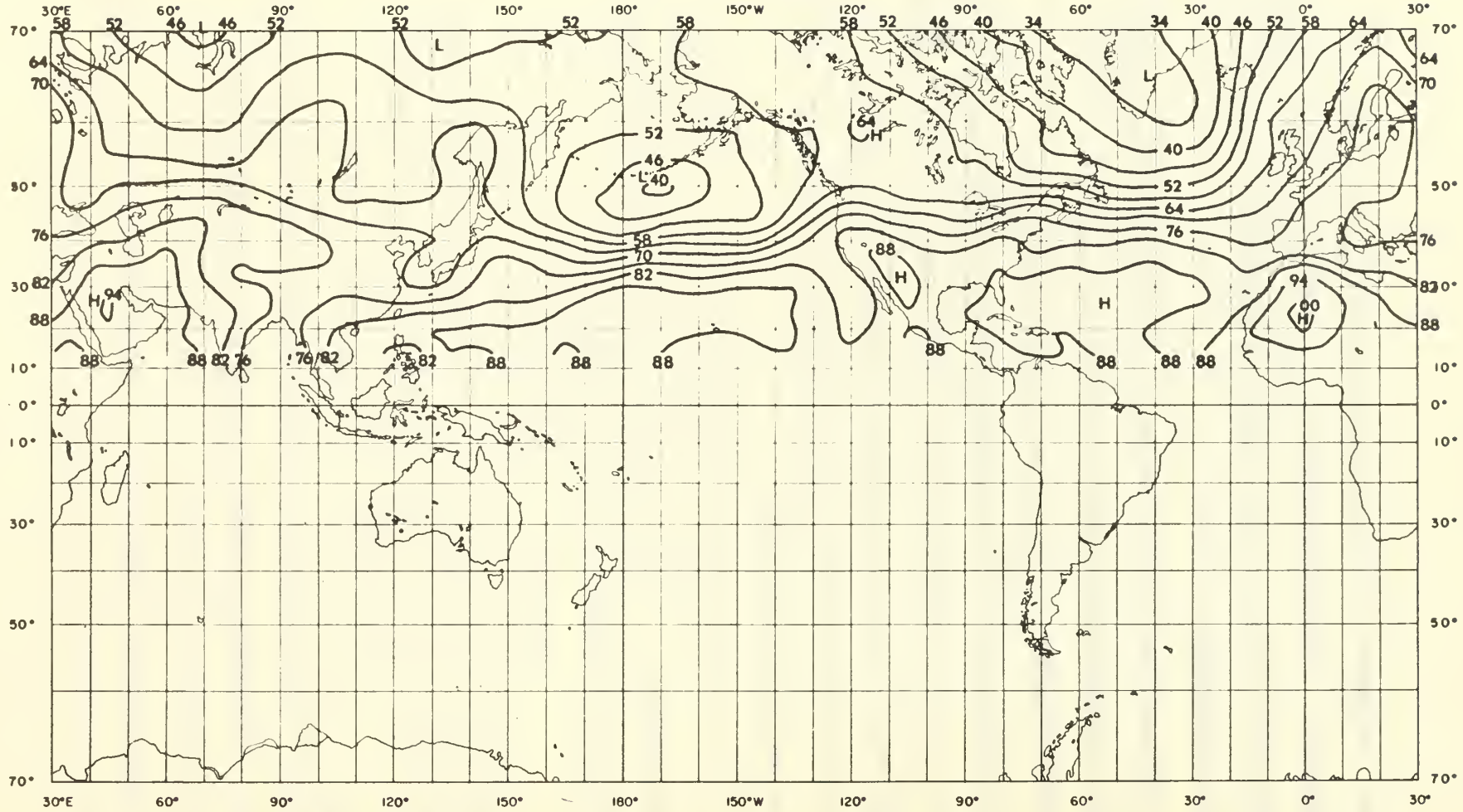

FIG. 3. The prognostic 5-day mean 500-mb geopotential field verifying at 00GMT, 14 June 1966 (i.e., t= 06).

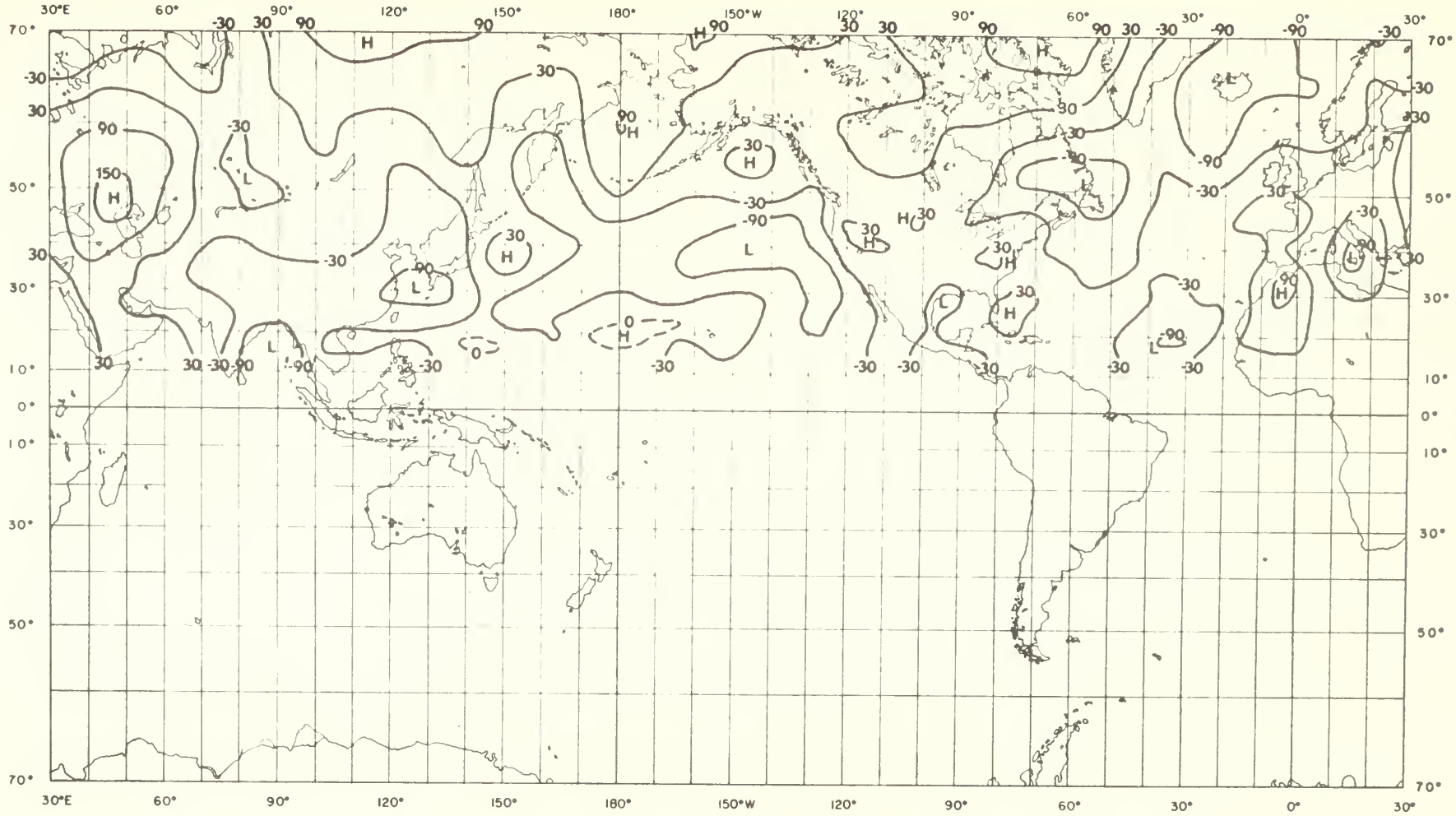

FIG. 4. The error-field (in gpm) at 00GMT, 14 June 1966, resulting from the prognosis in Fig. 3

corresponds to an R.M.S. error of 56.628 gpm. At t=12, the verifying field $\tilde{Z}_5(12)$ explained .8616 of the variance of $\tilde{Z}_5(12)$, with an R.M.S. error of 57.549 gpm. These two examples were selected to show that the prognostic update procedure outlined here seems to behave rather well throughout the time series.

Apart from the quoted R.M.S. standard errors in Z_5 for the two prognostic examples, the prognoses nevertheless give good definition of the verifying large-scale map features. A comparison of Fig. 3 with Fig. 2 gives rather good evidence of this statement. The geographic distribution of the actual troughs and ridges in the westerlies in Fig. 2 is clearly identifiable in the prognoses of Fig. 3. For example, the ridges over western North America, and over western Europe are clearly evident in both Fig. 2 and 3, as is the low in the Aleutian area. The rather complicated wave structure over Eurasia (Fig. 2) has been well recaptured in Fig. 3.

Comparison of Fig. 5 with Fig. 6 shows that the latter depicts most of the features of the verifying map at t=12. Among these are the tilted SW-NE ridge over western North America, and a second ridge extending S-N in over the Atlantic into the Greenland area. In Fig. 6 troughs over western Europe and eastern North America are closely comparable to similar features on the verifying map Fig. 5. The well defined ridge predicted over Western Siberia in Fig. 6 is likewise identified in Fig. 5. Fig. 6 shows considerable more trough-ridge amplitude over the Pacific; however, this is an area of the verifying analysis (Fig. 5) where much of the daily perturbation-amplitude

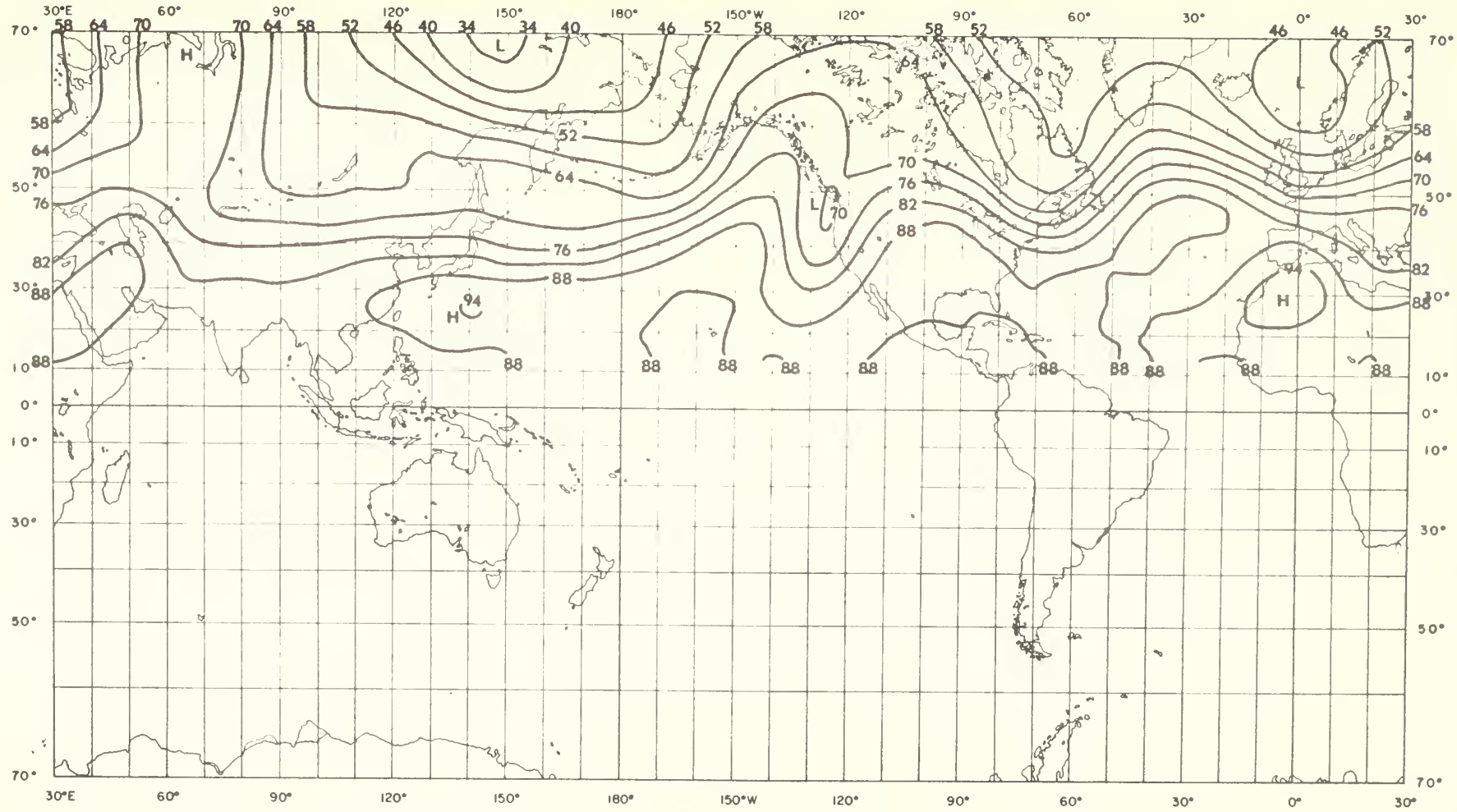

FIG. 5. The 5-day mean 500-mb analysis at 00GMT, 14 July 1966, (i.e., at t = 12) in Mercator coordinates.

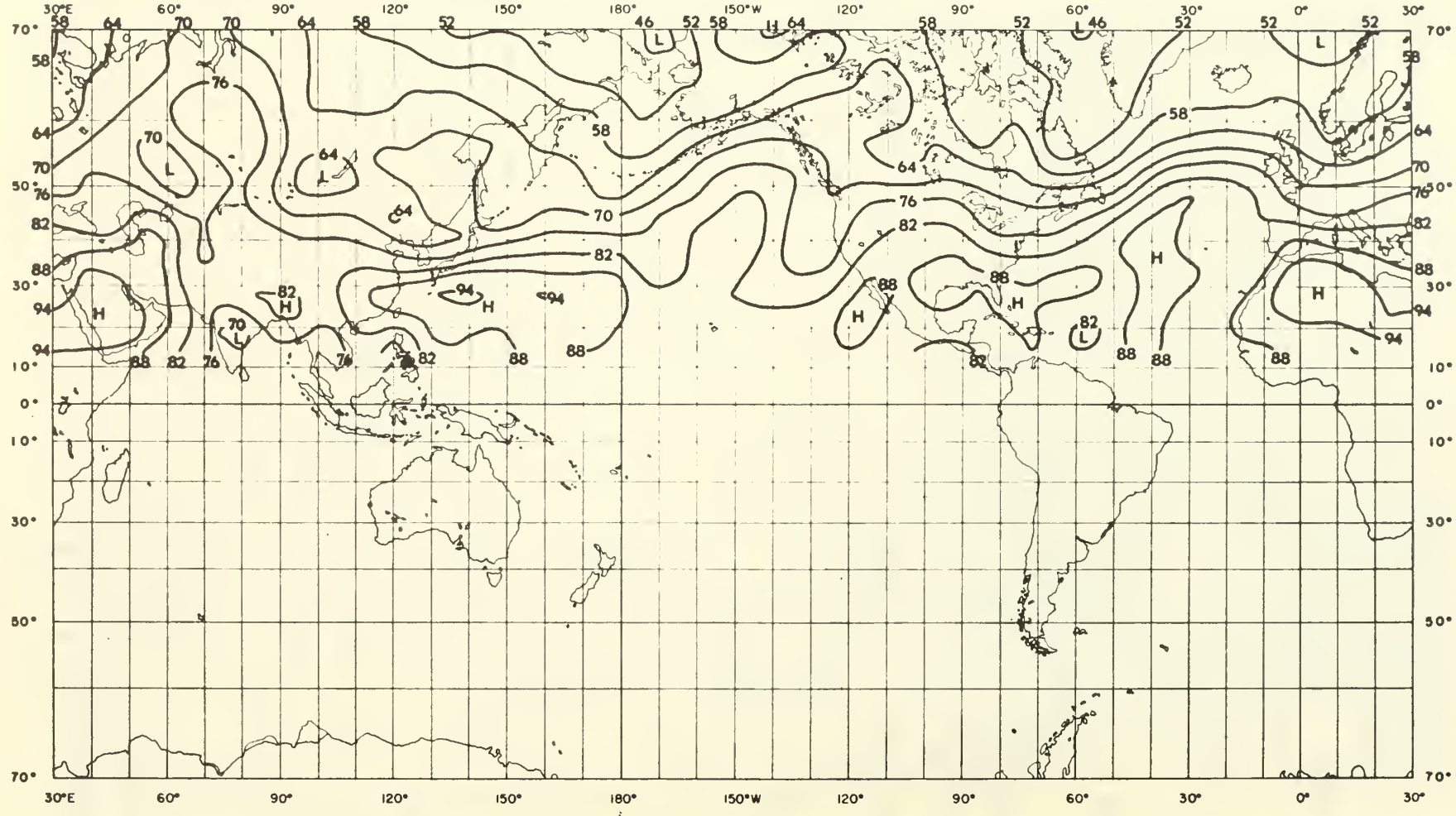

FIG. 6. The prognostic 5-day mean 500-mb geopotential field verifying at 00GMT, 14 July 1966.

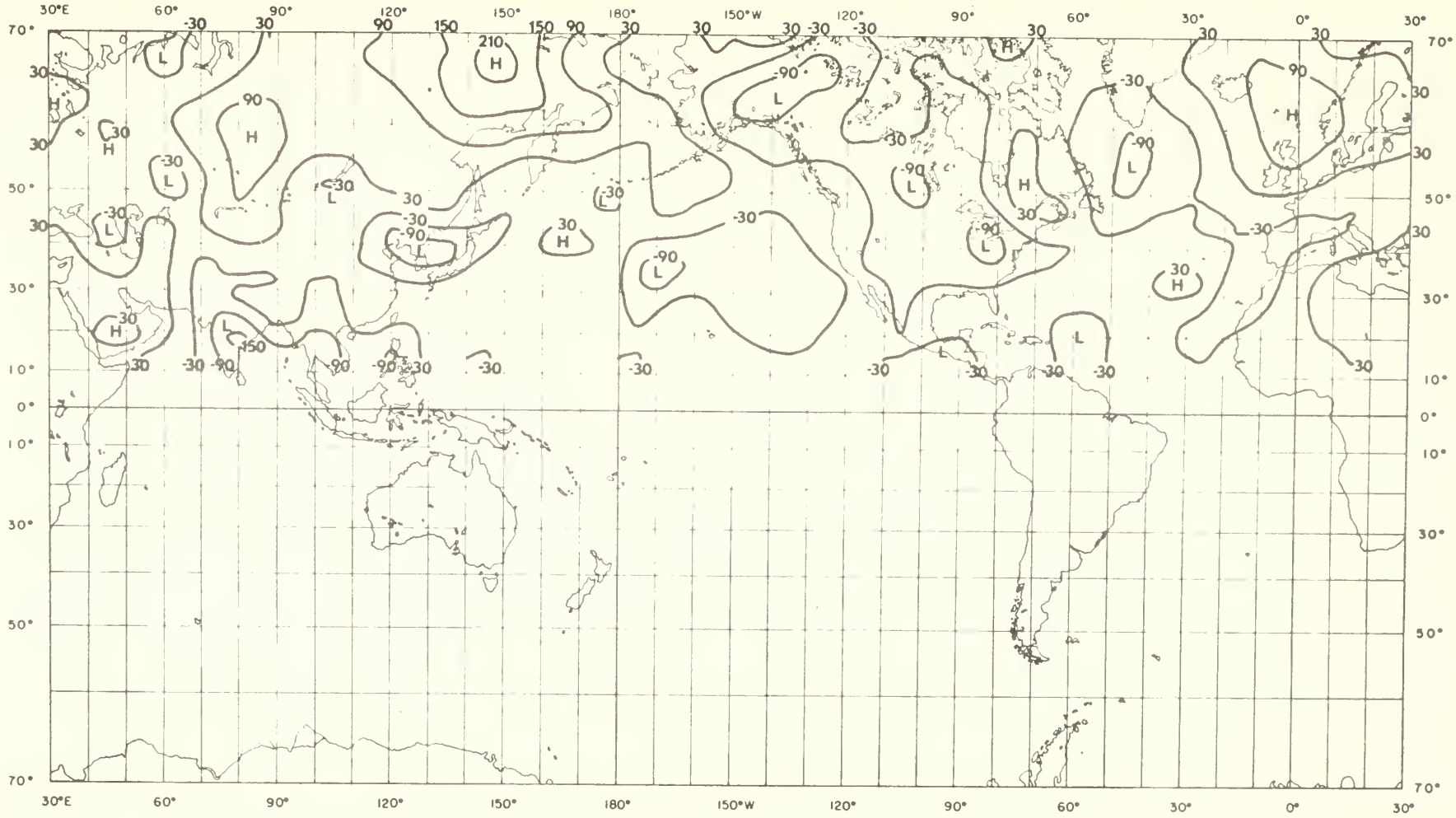

FIG. 7. The error-field (in gpm) at 00GMT, 14 June 1966, resulting from the prognosis in Fig. 6.

appears to have been smoothed out by the five-day time-averaging. The error chart, Fig. 7, has only a small proportion of its area covered by isopleths of error-magnitude in excess of 90 gpm, although an extreme value of +210 gpm is present in N.E. Siberia.

On the whole, however, both prognostic maps show features with good correlation to those of the verifying maps, where the latter have well-defined amplitudes.

Acknowledgments. My thanks are due to Mr. J. F. O'Connor of the Extended Forecast Section of the National Oceanic and Atmospheric Administration for use of the data tapes of the 500-mb fields used here. I am equally indebted to the National Space Science Data Center of the National Aeronautics and Space Administration for their kind assistance in providing the NIMBUS II composited MRIR data-charts. The programming assistance of Mr. Russell Schwanz is acknowledged, as is the data-processing and general assistance of the technical staff of the Meteorology Department, Naval Postgraduate School.

Appendix

TABLE 5. Coefficients of one-predictor (Test 1), three-predictor (Test 2), and five-predictor (Test 3) specification equations based upon data arrays at times t+1, (t+1 =2, ..., 15).

Analysis time	Test 1		Test 2				Test 3					
	A_5	A_o	B_5	B_4	B_3	B_o	C_5	C_4	C_3	C_2	C_1	C_o
t+1 = 2	.84981	103.498	.71293	4.37814	- .91172	- 747.153	.71172	9.25292	-4.10645	-16.85219	10.31656	341.258
3	.74554	191.819	.61303	3.64728	- .58259	- 555.778	.61475	7.15553	-3.08274	-13.30495	8.32237	344.890
4	.95328	29.686	.74731	5.48263	-1.10513	-1024.820	.74879	7.59886	-1.88581	- 8.60242	2.41488	72.099
5	.86682	99.287	.75076	5.29390	-1.96901	- 732.675	.74621	7.64621	-3.50324	- 8.00389	5.12921	-275.904
6	.93953	50.591	.84519	5.16481	-3.35899	- 381.161	.83924	7.73594	-5.38989	-11.34598	11.15528	-475,558
7	.95588	26.454	.91056	3.91661	-3.00754	- 195.120	.92139	6.09759	-4.87776	- 8.36942	7.13249	6.389
8	.82757	126.039	.76360	5.49224	-3.33576	- 419.548	.75405	6.87860	-4.14191	- 5.14454	3.36469	-150.026
9	.89269	84.986	.83884	3.11051	-1.79824	- 238.446	.82897	4.24500	-2.05523	- 4.33742	1.50659	199.500
10	.88652	86.714	.78319	4.54530	-2.45396	- 417.710	.75400	5.80199	-2.20904	- 5.57426	1.22915	220.237
11	.87318	99.676	.82413	3.94762	-2.64217	- 222.523	.80439	5.71786	-3.54875	- 6.08346	4.41213	- 49.778
12	.95895	32.602	.91456	1.93683	- .87692	- 225.597	.92672	4.14380	-2.45317	- 8.37135	5.66967	232.050
13	.85403	117.701	.76316	4.24528	-2.18038	- 382.744	.77043	5.94907	-3.89013	- 7.82676	7.96463	-416.462
14	.78965	161.026	.74356	7.44497	-5.68076	- 288.229	.71483	9.67274	-6.72892	- 8.63294	5.60993	127.480
15	.85041	99.767	.81310	3.88065	-2.60509	- 225.530	.82872	5.30726	-4.67465	- 6.19384	7.14397	-287.860

29

REFERENCES

Anderson, T. W., 1960: An introduction to multivariate statistical analysis. New York, John Wiley and Sons, 324 pp.

Dixon, W. J., 1966: Biomedical Computer Programs. Los Angeles, University of California Press, 600 pp.

Jensen, C. E., J. S. Winston, and V. R. Taylor, 1966: 500-mb Heights as a linear function of satellite infrared radiation data, Mon. Wea. Rev., 94, 641-649.

Haltiner, G. J., 1971: Numerical Weather Prediction. New York, John Wiley and Sons, 317 pp.

Martin, F. L., 1969: Statistical specification of the 500-mb height fields using smoothed medium-resolution radiometric fields of NIMBUS II. J. Appl. Meteor., 8, 668-686.

Miller, R. G., 1962: Statistical prediction by discrimination analysis. Meteor. Monogr. 4, No. 25, 54 pp.

NIMBUS II, User's Guide, 1966: National Space Science Data Center, Goddard Space Flight Center, NASA, Greenbelt, Md., 229 pp.

DISTRIBUTION LIST

Officer in Charge Naval Weapons Engineering Support Activity Detachment (FAMOS) 3737 Branch Avenue, Room 307 Hillcrest Heights, Maryland 20031	10
Officer-in-Charge Environmental Prediction Research Facility Naval Postgraduate School Monterey, California 93940	2
Commanding Officer Naval Weather Service Command 3101 Building 200 Washington Navy Yard Washington, D. C. 20390	1
Commanding Officer Fleet Numerical Weather Central Monterey, California 93940	1
Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
Dean of Research Administration, Code 023 Naval Postgraduate School Monterey, California 93940	2
Department of Meteorology Reference Center Naval Postgraduate School Monterey, California 93940	1
Professor F. L. Martin Department of Meteorology, Code 51 Naval Postgraduate School Monterey, California 93940	12
Defense Documentation Center (DDC) Cameron Station Alexandria, Virginia 22314	12

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Naval Postgraduate School
Monterey, California 93940

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

3. REPORT TITLE

Regression Methods in 5-Day Mean Height Prediction at 500mb Using Mean NIMBUS II Composited Infrared Data for Identical Periods in the Update Procedure

4. DESCRIPTIVE NOTES (Type of report and, inclusive dates)

Technical report, 31 March 1972

5. AUTHOR(S) (First name, middle initial, last name)

Frank L. Martin

6. REPORT DATE: 31 March 1972

7a. TOTAL NO. OF PAGES: 33

7b. NO. OF REFS: 7

8a. CONTRACT OR GRANT NO.

b. PROJECT NO.

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S): NPS-51MR720301A

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES

This research was supported by Foundation Research Program.

12. SPONSORING MILITARY ACTIVITY

Naval Postgraduate School
Monterey, California 93940

13. ABSTRACT

A series of 15 successive 5-day mean 500mb height fields in gridprint form covering the quasi-hemispheric area between 14.8 - 70.2N latitude was used in this experiment. The time period extended between 15 May 1966 through 28 July 1966. The stepwise regression procedure was employed to derive statistical estimator equations for each mean map field $Z_5(I,J,t+1)$ in terms of the preceding field data $Z_5(I,J,t)$ as the primary predictor, $(t+1 = 2, \ldots, 15)$. As additional predictors, the 5-day mean composited gridpoint values of the NIMBUS II MRIR equivalent black body temperatures in the water-vapor channel (denoted T_1) and the window channel (denoted T_2) were introduced for both analyses times t and t+1. The five-predictor regression equations thus developed for $Z_5(I,J,t+1)$ proved to have substantial statistical significance, both from the standpoint of the specification and of prediction of $Z_5(I,J,t+2)$ valid five days later $(t+2 = 3, \ldots, 15)$. In the latter connection, the prognosis of $Z_5(t+2)$ was significantly improved when all five predictors were included, as compared with the case when the data of either radiometric channel $T_1(I,J)$ or $T_2(I,J)$ at times t+1, t+2 was deleted from the prediction equations.

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
NIMBUS II						
Composited NIMBUS II Infrared data						
Equivalent blackbody temperature						
Five-day mean height-prediction						
Regression methods						
Verification on independent data						